

Webサービスへの 自然言語処理の適用

2011年3月2日

ニフティ株式会社

松井くにお

- 言葉は「つなげる」、「集める」、「換える」こと
によって付加価値がつく
- 「つなげる」とは
 - 情報抽出した結果をつなげて、新たな知識発見
- 「集める」とは
 - ウェブにあるソーシャルな情報を集め評判分析
- 「換える」とは
 - 言葉を翻訳して、海外への情報発信

今までの利用例

「表参道ヒルズ」「六本木ヒルズ」「丸ビル」の評判数の推移

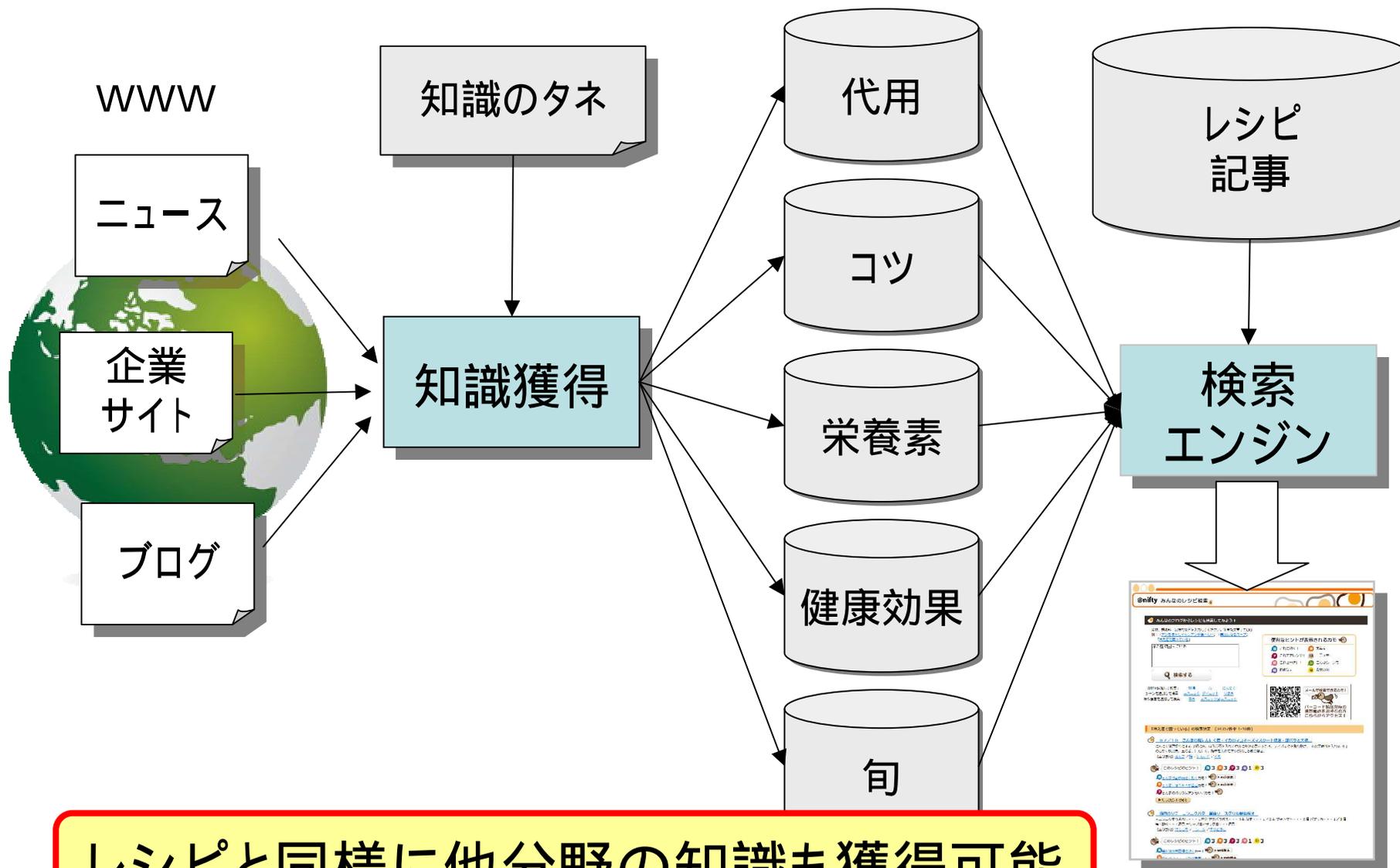


最近の利用例

「集めてつなげる」みんなのレシピ検索

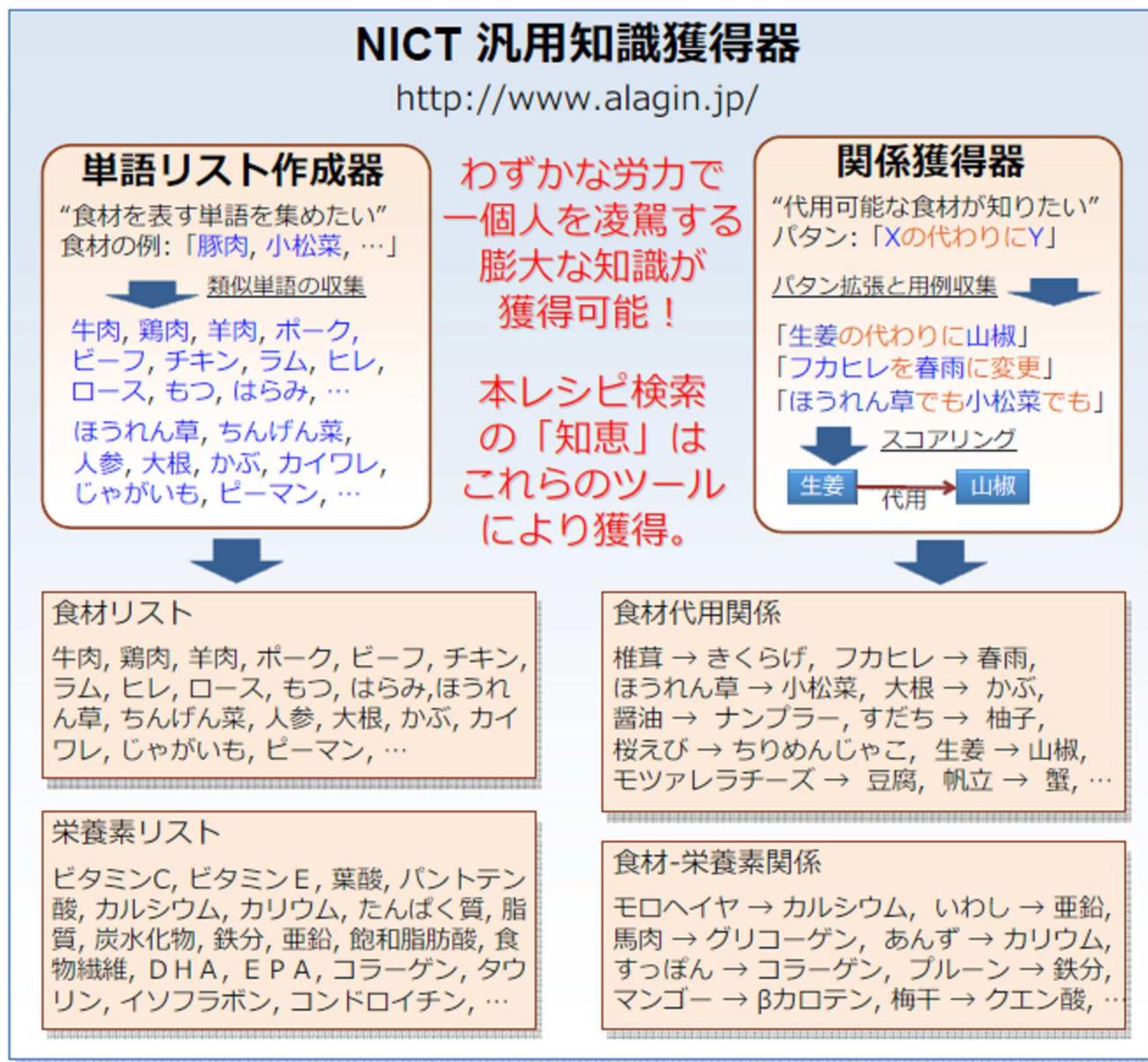


- 冷蔵庫の余り食材で検索
 - 食材検索は当たり前に出れます
- 「美肌になるスープ」「ビタミンCを摂れる料理」
 - 効能で検索
- 「アジが釣れたんだけど、イタリアンで、バジルを使った料理はある？」
 - 自然文検索
- 「疲れた時の夜食」
 - 気分から検索



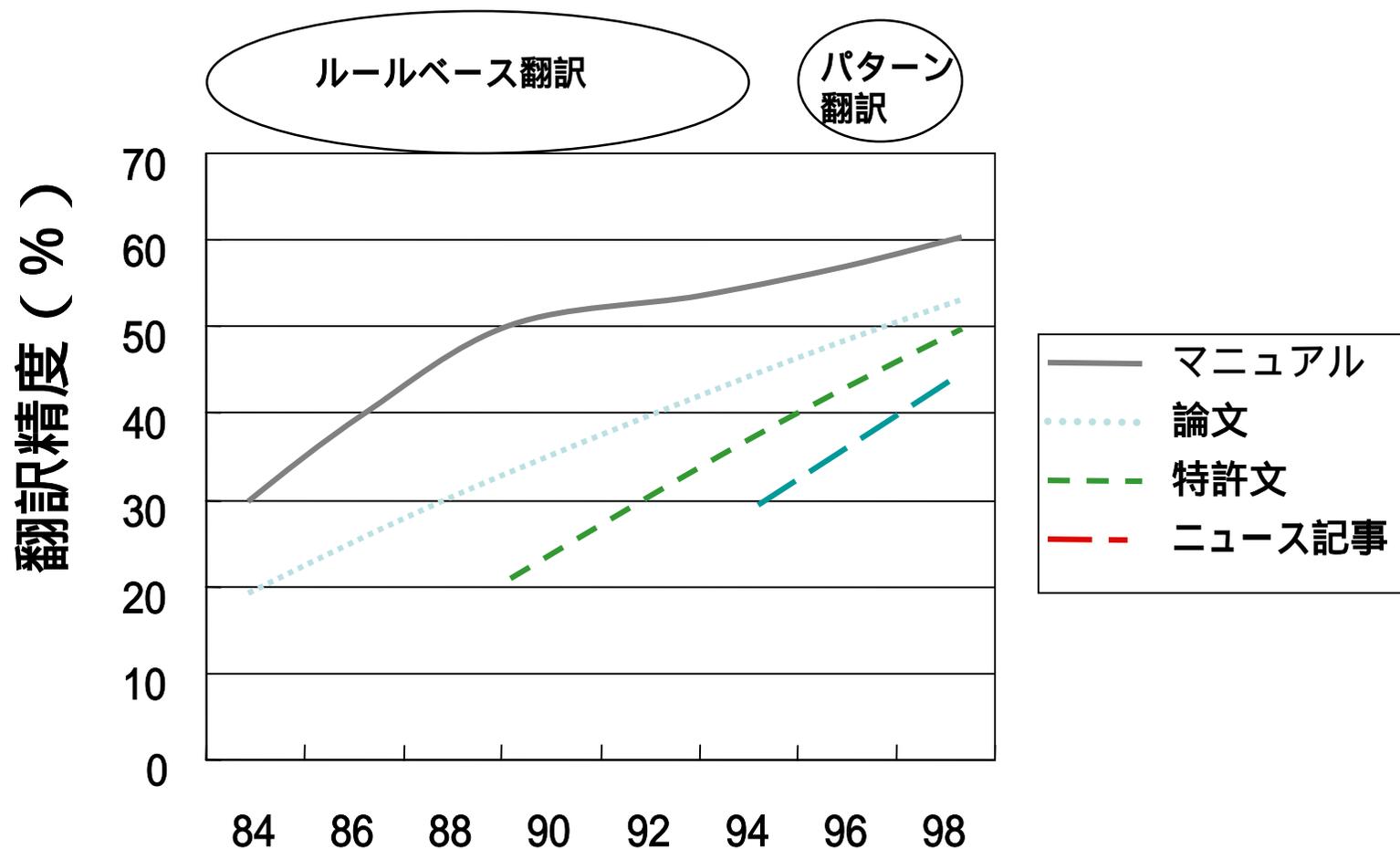
レシピと同様に他分野の知識も獲得可能

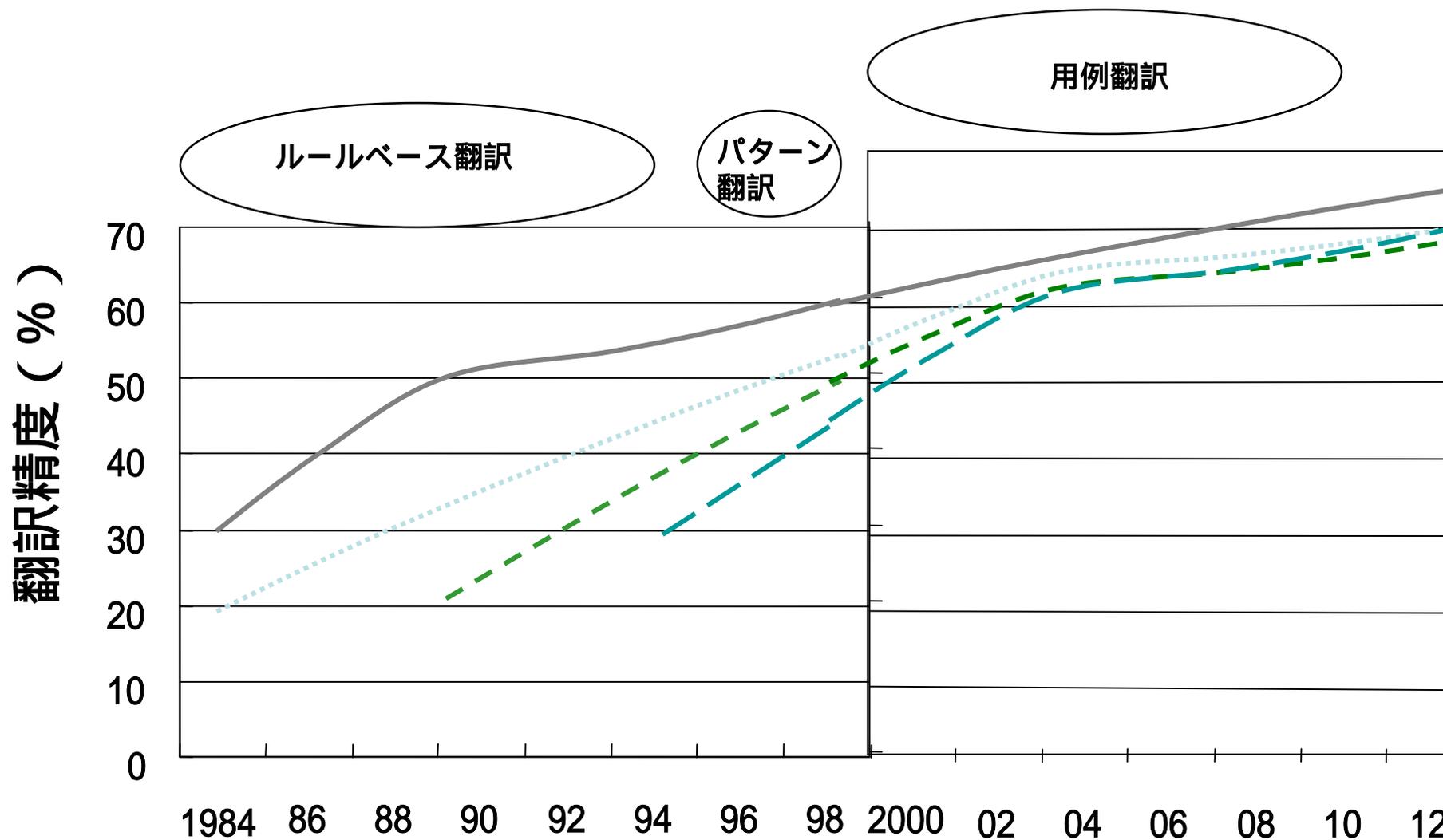
- 単語リスト生成
 - 特定のクラスの語句を集める。
 - 単語の係り受け関係の分布の類似度により、語句の間の類似度を求める。
- 関係獲得
 - 特定の関係を持つ2つの語句の対を集める。
 - 種となるデータとパターンを人手で与え、パターンと対を交互に取得する。



- レシピ検索
- 獲得した知識による検索機能の拡張
 - 関連する語句による検索
e.g. 冷え症 にんにく
- 検索をより柔軟に
 - ユーザが入力したクエリそのもの以外でも探せる。
(ex.)ファッション
 - 寒いとき用のレギンス
 - 旅行で着るワンピース
 - 結納に合うネクタイ
 - 着痩せするニット
 - しぼりこみ検索
(ex.)対話的なしぼりこみ
 - 「レギンス」と入れると
素材, 丈の長さ, 色, デテールなどをシステムが聞いてくる

- コンテンツ生成
 - 収集した知識をコンテンツとして利用
e.g. コツ
 - 味噌汁にラー油
 - 野菜ジュースにオリーブオイル
 - マッシュポテトにケチャップ
 - e.g. 名物
 - 北海道 バターサンド
 - 神戸 奈良漬
- ブログをデータベース化
 - ブログから特定の話題の記事だけを選別することにより、データベースとして用いる
e.g. レシピを含む記事 (2億記事 20万)





- 原文：講演概要から抜粋
 - インターネットの普及によって、「プロダクト」から「サービス」へと世の中のニーズが変わり、発信された言葉をそのまま読むだけでなく、言葉を「つなげる」、「集める」、「換える」ことによってその情報自体に付加価値を与えることができるようになった。
- @nifty翻訳で翻訳すると
 - Needs in "Service" from "Product" and the world changed by the spread of the Internet, and the sent word not only was read as it was but also the word was given and it came to be able to give the additional value to the information by "Tied", "Collect", and "Change it".
- google翻訳で翻訳すると
 - The spread of the Internet, "product" to "service" the world needs to change, not just reading the words as it was originated, the word "Common", "collect", "replace" that information by itself
Now you can add value.

浮かび上がった 問題点

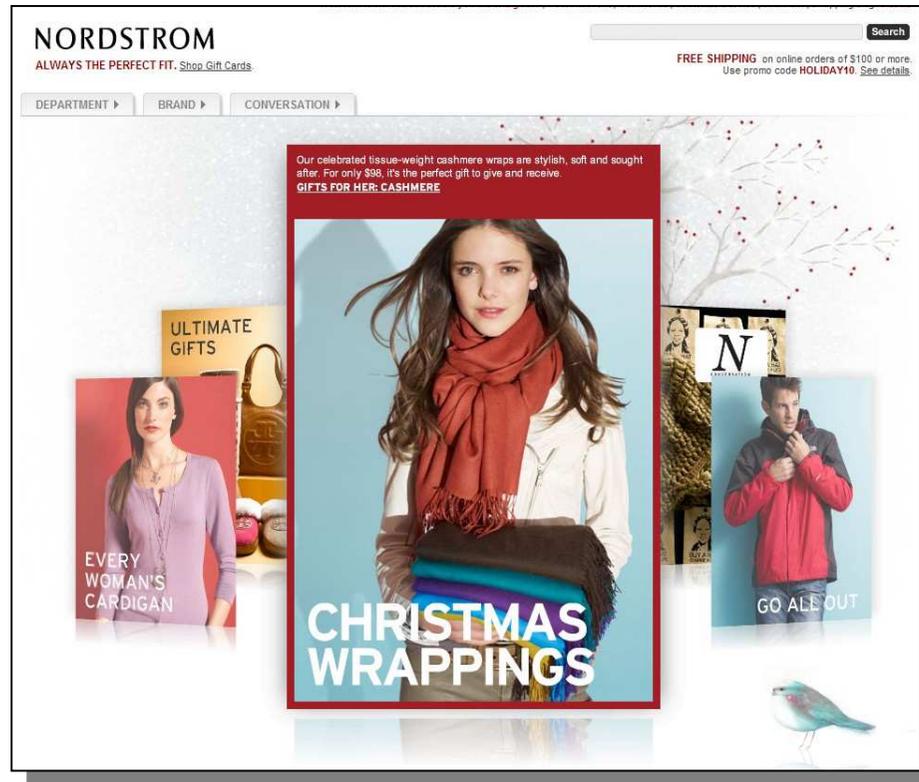
- 知識発見
 - どんな知識を発見するかは人間の周到な準備が必要
- 評判分析
 - 有効な評判(良くも悪くも)は人間が解釈して対応
- 機械翻訳
 - 情報発信のためにはポストエディティングが必要

コストパフォーマンスが悪い

新たな展開

- Advocacy marketing
 - 最初は不利な情報に見えても消費者の信頼を得て、企業業績を向上させるマーケティング
- Open innovation
 - 一企業だけでは処理できない課題を外部の力を得て解決する
- ライフログ活用
 - サービス価値向上、マーケティング情報としての利用

ノードストロームの例

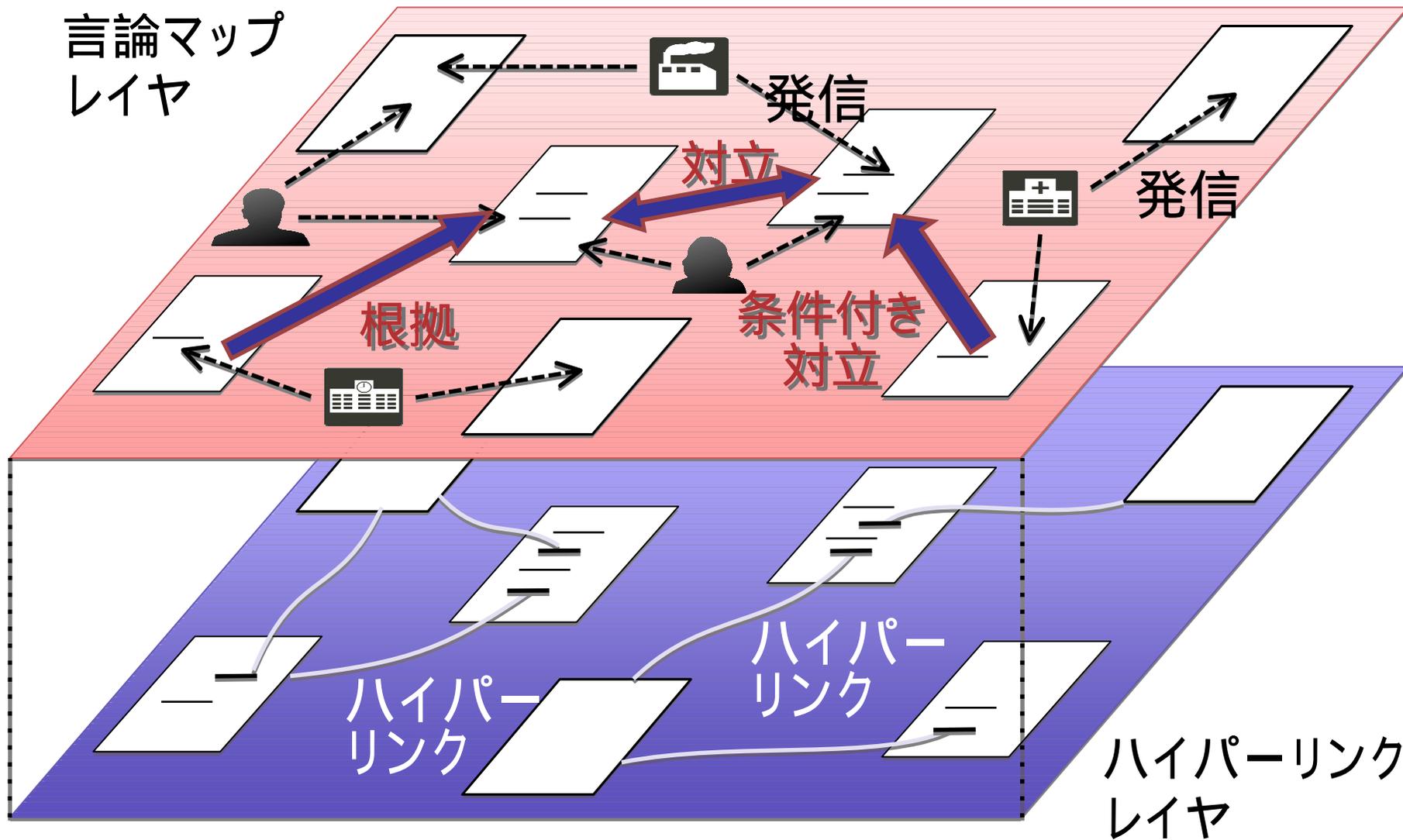


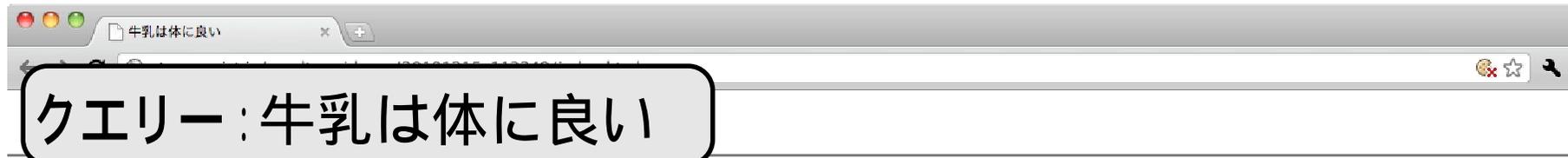
顧客が欲しい商品が在庫として無い場合、他店の在庫を調べて案内したという逸話(都市伝説?)

データ、言語処理、リアル店舗の統合利用

- 個人が力を持ち始める
 - 消費者の主役としての個人
 - インターネットの普及による個人による情報発信
- 企業は顧客との“信頼”が重要に
 - クチコミの重要性
 - 在庫が無い、不良品情報などを隠さない

これこそが「Advocacy marketing」





賛成根拠
(18文)

牛乳には良質のたんぱく質、ビタミンA・ビタミンB・ビタミンE・ビタミンDもバランスよく含んでいるため、身体全体の調子を整え、体力向上に役立ちます^{log1}

from 牛乳

反対根拠
(12文)

だから牛乳は消化不良食品になりますから身体に合わない判断したら飲まない方がよい^{log1}

from 良い宇宙人と悪い宇宙人

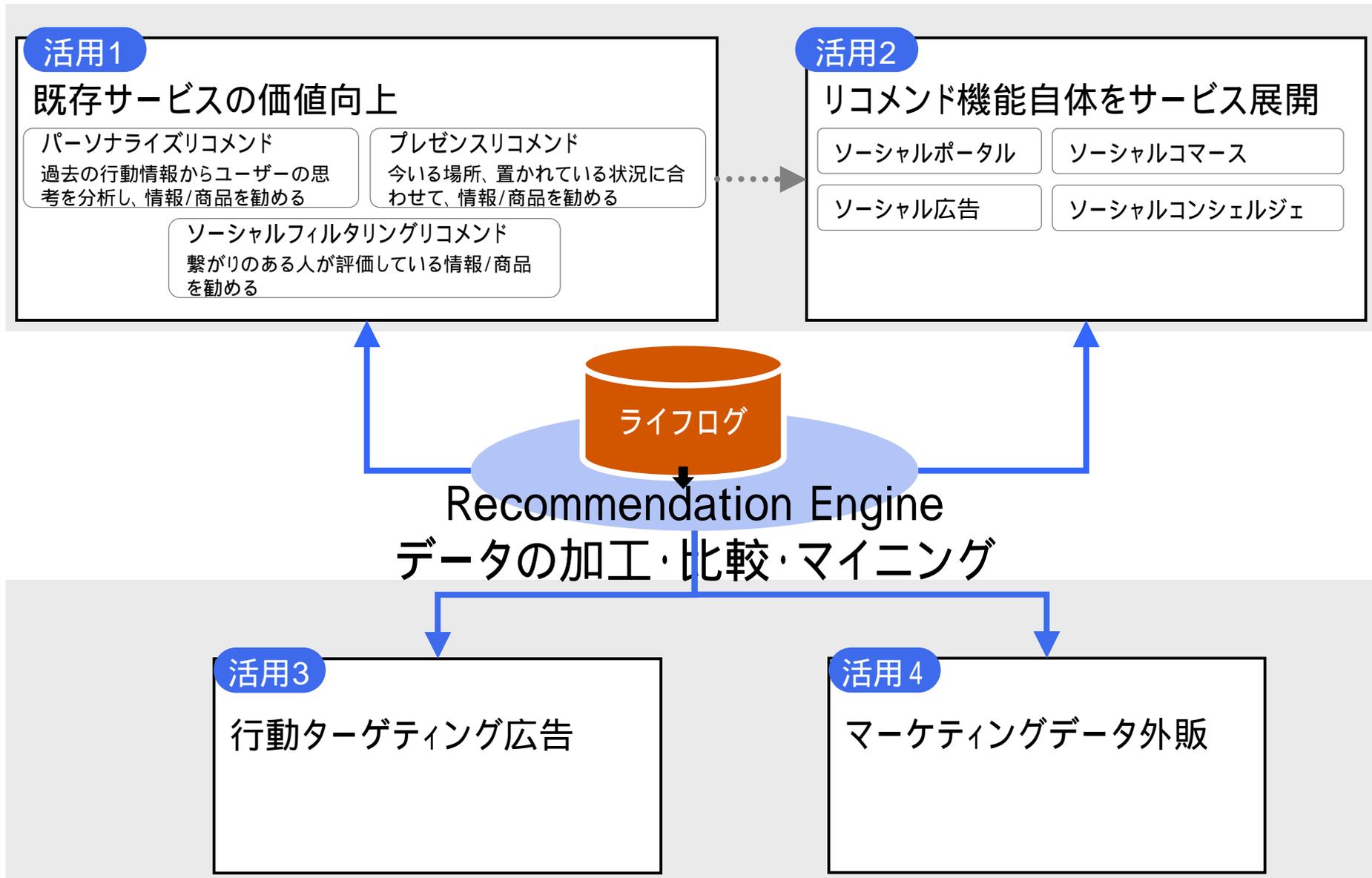
- 牛乳は体質によっては体に悪いので、無理強いは良くないと思いません^{log1}

賛成根拠: 牛乳には良質のたんぱく質、ビタミンA・ビタミンB・ビタミンE・ビタミンDもバランスよく含んでいるため、身体全体の調子を整え、体力向上に役立ちます

- 一般的に行なわれている殺菌法とちがい、牛乳全体に加わる熱が少ないため、新鮮な生乳のすっきりとしたキレのある味を守ります^{log1}
from 森永乳業株式会社【ニュースリリース】
- しかし、悪影響を及ぼす光の波長をブロックする脂肪のレベルが高いため、牛乳全体には影響はほとんどない^{log1}
from 瓶の牛乳は蛍光灯の影響を受けて酸化臭がするようにな...

- 一般によく売られている牛乳は高温殺菌しているので、体に役立つ菌まで殺菌してしまいます^{log1}
from 「わが村は美しく-北海道」運動2004
- 牛乳もヨーグルトもチーズも、それ自体がエネルギーを持っているから、たくさん食べ過ぎるともちろん太ります^{log1}
from -毎日3回、からだにいいコト-3-A-Day

反対根拠: 一般によく売られている牛乳は高温殺菌しているので、体に役立つ菌まで殺菌してしまいます



- Webサービスとして言語処理を活用したい
 - But 人間の介在が必要でコストがかかる
 - 精度向上は産官学の連携が必須
- ソーシャルメディアの情報を活用したい
 - But ユーザの表現に制約を課すのは無理
 - 産業日本語への自動変換に期待
- 新しいビジネス領域を開拓したい
 - 言語処理、データ処理、リアル連携を一体で